Packet-loss modeling with state duration constraints and VoIP based on perceptual quality maximization

Nestor Becerra Yoma, Carlos Busso, Juan Inzunza and Fernando Hueinupan
Electrical Engineering Department, University of Chile
Av. Tupper 2007, P.O.Box 412-3, Santiago, CHILE
nbecerra@cec.uchile.cl

Abstract

In this paper a Gilbert-gamma topology is proposed to model packet-loss processes in TCP/IP connections. The proposed topology introduces state duration modeling with gamma distributions. When compared to the ordinary Gilbert model, the proposed topology substantially improves the likelihood of observed packet-loss processes, and gives reductions as high as 70% in the subjective quality estimation of speech transmitted over TCP/IP networks. The ordinary Gilbert topology can give an error in bit rate as high as 100% in the context of VoIP based on perceptual quality maximization. This error can dramatically be reduced with the approach presented here.

1. Introduction

As it is well known, TCP/IP networks introduce a packet-loss process. The accurate description of this process is very relevant to estimate the perceptive quality of real-time applications (e.g. VoIP, audio and video streaming), to evaluate speech recognition systems with simulated VoIP conditions, and to evaluate the response of coder schemes to packet-loss. In [Bolot, 1993] the packet-loss process in the Internet was studied and modeled as a two-state Markov topology or Gilbert model (Fig. 1). This two-state Markov topology has been employed in several papers to model the packet-loss process to analyze the TCP protocol [Abouzeid et. al, 2000], the response of FEC (Forward Error Correction) [Bolot et. al, 1999] [Frossard, 2001], and to evaluate speech recognition algorithms [Peláez-Moreno et. al, 2001] [Quercia et. al, 2002]. However, according to [Zhu et. al, 1995], where a two-dimensional discrete-time Markov chain was proposed to model the time interval between two consecutive bursts of lost packets, the Gilbert topology does not model properly the temporal behavior of a non-blocking period (i.e. burst of successfully transmitted packets). On the other hand, the prediction of the subjective quality of speech transmitted over TCP/IP employing network parameters has been addressed in [Clark, 2001], where a four-state Markov chain was presented to model bursts of lost packets, and in [Mohamed et al., 2001] and [Sun & Ifeachor, 2002], where neural-network based approaches were proposed.

In the ordinary Markov topology, the probability transitions are defined by constants that lead to geometric state duration distributions. However, observed bursts of lost packets in TCP/IP connections show that this model is inaccurate in many cases. The contribution of this paper concerns a Gilbert topology that incorporates state duration modeling with gamma functions (Gilbert-gamma). When compared to the ordinary Gilbert model, the proposed topology substantially improves the likelihood of observed packet-loss processes, and dramatically increases the accuracy of the subjective quality estimation of speech transmitted over TCP/IP networks. Finally, the results presented here have not been found in the specialized literature and can be employed in other real-time applications such as audio and video streaming.

2. The Gilbert Model and State Duration Distribution

The packet-loss process is modeled here by the packet-loss rate (PL) and the probability distribution of burst length (*BL*), which in turn corresponds to the number of consecutive lost or transmitted packets. The basic Gilbert model uses state 0 to represent a packet that was lost and state 1 for a packet that successfully reached the destination (Fig. 1). The probability that *BL* consecutive packets are lost is geometrically distributed and equal to,

$$Pr(\text{burst length} = BL \text{ in state } 0) = (a_{0,0})^{BL} \cdot a_{0,1} \quad (1)$$

where

$$a_{1,0} = \frac{PL}{E_0[BL] \cdot (1-PL)} \quad (2)$$

$$a_{0,1} = \frac{1}{E_0[BL]} \quad (3)$$

$$Var_0[BL] = \frac{a_{0,1} - 2}{(a_{0,1})^2} \quad (4)$$

where is the burst length in packets; $E_0[BL]$ and $Var_0[BL]$ denote the expected value and variance, respectively, in terms of lost packet (i.e., for state 0). Note that the Gilbert model is completely defined by two variables, $a_{0,1}$ and $a_{1,0}$, and that these variables can be estimated with PL and $E_0[BL]$, which in turn define $Var_0[BL]$.

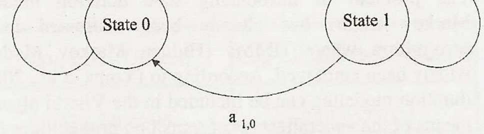

Figure 1: The Gilbert model

The evaluation of the ordinary Gilbert model was done using SoLPs (SoLP-Sequence of Lost Packets) that correspond to time sequences that indicate which packets are successfully transmitted and which are lost. These sequences describe a TCP/IP connection from the packet-loss point of view. The following methodology was implemented: UDP packets were transmitted at a constant bit rate over two types of real Internet connection with a client process at UCh (Universidad of Chile) in Santiago, Chile, either with a 100 Mbps Fast Ethernet or with a 256 Kbps cable/DSL access. In both cases the server process was implemented in a host at UNM (University of New Mexico) in Albuquerque, USA. For each type of access at UCh, twenty 20-minute connections were made and one SoLP per connection was recorded. Altogether, 2 types of access x 20 connections = 40 SoLPs were stored. The parameters PL and $E_0[BL]$ were estimated every 30-second frame from the SoLPs, and the distribution of in packet-loss bursts was evaluated in each 20-minute connection. As mentioned above, the transition probability is represented by a constant in the basic Gilbert model which gives a geometric probability distribution for state duration. As a consequence, the probability distribution of in packet-loss bursts should also be geometric but this model is not accurate in many cases where the distribution of is better fitted by a gamma function [12]. As a result, the Gilbert model could be improved by introducing state duration constraints to model more accurately the packet-loss process. Note that the geometric and gamma distributions are similarly monotonic when the expected value of is close to 1.

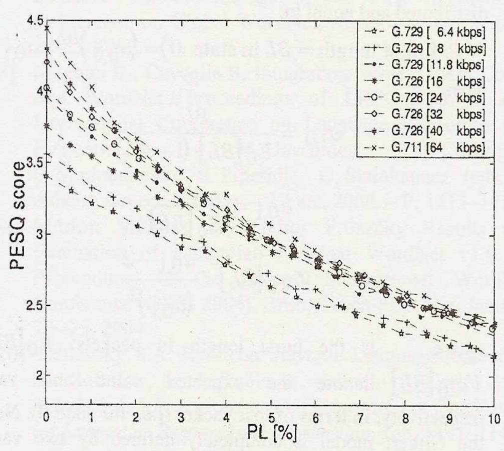

Figure 2: Example of PESQ vs. P curves estimated with the ordinary Gilbert model (E[BL] =3).

3. Including Temporal Restrictions in the Gilbert Model

The problem of introducing state duration modeling in Markov chains has already been addressed in speech recognition where HMMs (Hidden Markov Models) has widely been employed. According to [Yoma et al., 2000] state duration modeling can be included in the Viterbi algorithm by means of the generalization of transition probabilities [12]:

$$a_{i,j}^{BL} = \text{Prob}(s_{t+1} = j \mid s_t = s_{t-1} = \ldots = s_{t-BL+1} = i) \quad (5)$$

where BL is the number of consecutive packets in state i up to time t; $j=i$, or $j=i+1$ (if $i=0$) or $j=i-1$ (if $i=1$) given the topology shown in Fig.1. Using these definitions for the transition probabilities, $a_{i,i}^{BL}$ and $a_{i,j}^{BL}$ can be estimated by:

$$a_{i,i}^{BL} = \frac{D_i(BL) - d_i(BL)}{D_i(BL)} \quad (6)$$

$$a_{i,i+1}^{BL} = \frac{d_i(BL)}{D_i(BL)} \quad (7)$$

where $d_i(BL)$ is the probability of state duration equal to BL packets and $D_i(BL)$ is the probability of state i being active for $t \geq BL$:

$$D_i(BL) = \sum_{t=BL}^{\infty} d_i(t) \quad (8)$$

To include the possible *min* and *max* durations, $min_i(BL)$ and $max_i(BL)$, respectively, the transition probabilities were modified to:

$$a_{i,i}^{BL} = \begin{cases} 1 & \text{if} \quad BL < min_i(BL) \\ 0 & \text{if} \quad BL > max_i(BL) \\ \dfrac{D_i(BL) - d_i(BL)}{D_i(BL)} & \textit{otherwise} \end{cases} \quad (9)$$

$$a_{i,i+1}^{BL} = \begin{cases} 0 & \text{if} \quad BL < min_i(BL) \\ 1 & \text{if} \quad BL > max_i(BL) \\ \dfrac{d_i(BL)}{D_i(BL)} & \textit{otherwise} \end{cases} \quad (10)$$

where $_i(BL)$ is modeled with the discrete gamma distribution given by:

$$d_i(BL) = K_i \cdot e^{-\alpha_i \cdot BL} \cdot BL^{p_i - 1} \quad (11)$$

where $BL = 1,2,3...$ is the duration of a given state i in number of packets, $\alpha_i > 0$, $p_i > 0$ and K_i is a normalizing term. The mean, $E_i[BL]$, the variance, $Var_i[BL]$, and the *min* and *max* durations were computed for state $i=0$ and $i=1$ by means of directly observing the bursts of lost (state 0) and transmitted (state 1) packets. The parameters α_i and p_i were estimated by:

$$\alpha_i = \frac{E_i[BL]}{Var_i[BL]} \quad (12)$$

$$p_i = \frac{E_i^2[BL]}{Var_i[BL]} \quad (13)$$

Note that the Gilbert-gamma model used statistics of bursts for both lost and successfully transmitted packets. In contrast, the basic Gilbert model used only and the expected value of in packet-loss bursts. It can easily be shown that, according to the Gilbert-gamma model, the packet-loss rate, , is given by:

$$PL \approx \frac{E_0[BL]}{E_0[BL] + E_1[BL]} \quad (14)$$

where $E_0[BL]$ and $E_1[BL]$ are the expected value of in lost and transmitted packet bursts, respectively.

4. The Gilbert model as a generator of SoLP

Each SoLP can be decomposed as a sequence of frames. The w^{th} frame in the th SoLP is defined as $SoLP_{n,w} = (SoLP_{n,w,1}; SoLP_{n,w,2}; \ldots; SoLP_{n,w,l}; \ldots; SoLP_{n,w,L})$ where is the length of frame in packets, and $SoLP_{n,w,l} = 0$ or $SoLP_{n,w,l} = 1$ if packet l was lost or transmitted, respectively. A Gilbert model, as any Markov chain, could be seen as a process generator. As a consequence, the log-likelihood of a SoLP given a Gilbert model can be estimated.

5. Results

The introduction of temporal restrictions in the Gilbert model was evaluated by computing the probability to generate SoLPs and by estimating the PESQ (Perceptual Evaluation of Speech Quality) as shown in Fig. 3. The 40 20-minute SoLPs described in section II were employed. In the first case the following procedure was adopted: , $E_i[BL]$, $Var_i[BL]$, and $\min_i(\tau)$ and $\max_i(\tau)$ in (9) and (10) were estimated every 30-sec window in each SoLP; second, the probability transitions $a_{0,1}$ and $a_{1,0}$ in the ordinary Gilbert model were computed according to (2) and (3); third, the parameters α_i and p_i of the gamma functions that model the burst of transmitted and lost packets were computed with (12) and (13); and finally, the log-likelihoods of each SoLP given the ordinary Gilbert and Gilbert-gamma models were evaluated with (16) and (17), respectively. The results are presented in Table 1 that corresponds to the average log-likelihoods within the set of 40 SoLPs. According to Table 1, the Gilbert-gamma model provides a significant increase in log-likelihood when compared to the ordinary Gilbert topology. This result strongly suggests that the introduction of state duration modeling with gamma distribution significantly improves the accuracy of the Gilbert model. Observe that the distribution of in transmitted packet bursts is also modeled with a gamma function. Moreover, the truncation of the conditional transition probabilities according to (9) and (10) also contributes to improve the description of the distributions of lost and transmitted bursts.

The Gilbert-gamma was also evaluated according to its ability to map the packet-loss process to the PESQ score. Speech samples from 400 utterances (40 speakers and 10 utterances per speaker) from LATINO-40 database (Linguistic Data Consortium, LDC) were employed. PESQ is described in the standard ITU-T P.862 and attempts to give a subjective evaluation of the degradation in speech quality as a result of the coding-decoding distortion and/or of the packet-loss. Three coders were considered: 64 kbps PCM G.711; 32 kbps ADPCM G.726; and, 8 kbps CS-ACELP G.729. The procedure presented in Fig. 3 is described as follows: first, each observed SoLP was employed to simulate a real TCP/IP, by discarding lost packets after coding, and then the PESQ score was evaluated after decoding; second, the same procedure was repeated by replacing the original SoLP with a sequence generated with the corresponding Gilbert or Gilbert-gamma models; finally, the error in PESQ score was computed. The Gilbert and Gilbert-gamma models were used as process generators to replace the original SoLP as mentioned above: , $E_i[BL]$, $Var_i[BL]$, $\min_i(\tau)$ and $\max_i(\tau)$, and α_i and p_i of the gamma functions (12)(13) were all estimated every 30-sec window in each SoLP; as a consequence, each real SoLP was represented by a sequence of Gilbert and a sequence of Gilbert-gamma models, and these sequences were employed to generate one SoLPs each. The average errors in PESQ score computed within the 40 SoLPs are shown in Table 2.

According to Table 2, the average errors in PESQ score are 3.83%, 3.37% and 3.10% with the 64 kbps PCM G.711, 32 kbps ADPCM G.726 and 8 kbps CS-ACELP G.729 coders, respectively, when the ordinary Gilbert model was employed to generate the SoLPs. These errors are reduced to 1.08%, 0.96% and 1.12% with the 64 kbps PCM G.711, 32 kbps ADPCM G.726 and 8 kbps CS-ACELP G.729 coders, respectively, when the ordinary Gilbert model is replaced with the Gilbert-gamma topology to generate the SoLPs. Consequently, the Gilbert-gamma model gave an overall reduction of 70% in PESQ error. This result strongly suggests that the Gilbert-gamma model describes a packet-loss process more precisely than the ordinary Gilbert model to reproduce the PESQ score given by the real SoLP. In other words, the proposed Gilbert-gamma topology dramatically improves the accuracy of prediction of a PESQ score from a packet-loss process, if the latter is represented by a model and its set of parameters.

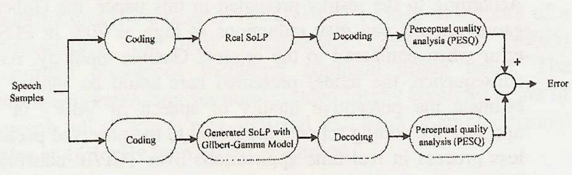

Figure 3: Scheme employed to compare the Gilbert and Gilbert-gamma models with respect to the accuracy to map a packet-loss process to the perceptual quality of speech transmitted over a TCP/IP network.

Table 1: Averaged log-likelihoods within the set of 40 SoLPs.

Log-Likel[Gilbert]	Log-Likel[Gilbert-Gamma]
-120,5466806	-118,0527024

Table 2: Error (%) between the PESQ score achieved with the real SoLP and the one obtained with the SoLPs generated by the Gilbert and the Gilbert-gamma models.

Coder	Gilbert Model Error [%]	Gilbert-Gamma Model Error [%]
G.729 (8 kbps)	3,0968	1,1193
G.726 (32 kbps)	3,3654	0,9583
G.711 (64 kbps)	3,8236	1,0782
Average	**3,4286**	**1,0519**

Finally, as shown in Fig. 2, the ordinary Gilbert model can give an error as high as 100% in the bit error if the coder is chosen based in the maximization of PESQ given a packet-loss process. This error could be reduced to 30 or 20% with the approach presented here.

6. Conclusions

The Gilbert-gamma model proposed in this paper attempts to overcome the fact that the transition probability is represented by a constant in the ordinary two-state Markov model, which in turn leads to a geometric probability distribution for state duration. This model is not accurate enough to describe bursts of lost-packets observed in TCP/IP connections. The method proposed here models the state duration distributions with gamma functions and introduces conditional transition probabilities that are truncated according to the minimum and maximum observed burst lengths in lost or successfully transmitted packet bursts. The results presented show that the proposed Gilbert-gamma model substantially increases the likelihood of observed SoLP when compared to the ordinary Gilbert topology.

As far as the ability to map packet-loss processes to PESQ scores is concerned, the Gilbert-gamma topology presented here dramatically improves the accuracy of the mapping from a packet-loss process to a PESQ score. According to the results presented in this paper, the Gilbert-gamma model can give reductions as high as 70% in PESQ error when compared to the original Gamma topology. As a consequence, the model presented here could be applied to estimate the perceptive quality of speech in VoIP or to optimize the coding scheme according to the observed packet-loss process in real-time applications over TCP/IP networks. The evaluation of speech recognition systems with an artificially introduced packet-loss process to simulate VoIP conditions should also benefit from the results discussed in this paper. Actually, the reduction in PESQ error reported here can provided reductions as high as 70% or 80% in the error of bit rate. Finally, the application of the Gilbert-gamma model to estimate the perceptive quality of audio and video streaming in TCP/IP networks from the observed packet-loss process is proposed as a future work.

7. Acknowledgments

The authors also wish to thank Prof. Ramiro Jordan and John Salas, from the University of New Mexico, USA, for having provided the remote host employed in the experiments reported here. The research described here was supported by Conicyt-Chile.

8. References

[1] Abouzeid, A.A., Roy, S., Azizoglu, M. (2000). Stochastic modeling of TCP over lossy links. INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE , Volume: 3 , 26-30 Mar 2000 Page(s): 1724 -1733 vol.3

[2] Bolot, J. C. (1993). Characterizing end-to-end packet delay and loss in the Internet. Journal of High Speed Networks, 2(3):305--323, 1993.

[3] Bolot, J.-C., Fosse-Parisis, S., Towsley, D.(1999). Adaptive FEC-based error control for Internet telephony. INFOCOM '99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE , Volume: 3 , 1999.

[4] Clark, A. D. (2001). Modeling the Effects of Burst Packet Loss and Recency on Subjective Voice Quality. IPTEL'2001, pp. 123-127, April, New York, 2001.

[5] Frossard, P. (2001). FEC performance in multimedia streaming. IEEE Communications Letters, Volume: 5 Issue: 3, Mar 2001. Page(s): 122 -124.

[6] Mohamed, S., Cervantes-Perez, F., Afifi, H.(2001). Integrating networks measurements and speech quality subjective scores for control purposes. INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE , Volume: 2 , 2001. Page(s): 641 -649 vol.2

[7] Peláez-Moreno, C., Gallardo-Antolín ,A., Díaz-de-María,F. (2001). Recognizing Voice over IP: a robust front-end for speech recognition on the World Wide Web. IEEE Transactions on Multimedia, Vol. 3, No.2, pp.209-218, June 2001.

[8] Quercia, D., Docio-Fernandez, L., Garcia-Mateo, C., Farinetti, L., De, C.(2002). Performance analysis of distributed speech recognition over IP networks on the AURORA database. Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). IEEE International Conference on , Volume: 4 , 2002.Page(s): IV-3820 -IV-3823 vol.4.

[9] Sun, L., Ifeachor, E.C.(2002). Perceived speech quality prediction for voice over IP-based networks. Communications, 2002. ICC 2002. IEEE International Conference on, Volume: 4, 2002 Page(s): 2573 -2577 vol.4.

[10] Yoma, Nestor B., McInnes, F., Jack, M., Stump, S., Ling, L. (2000). On including temporal constraints in viterbi alignment for speech recognition in noise. IEEE Transactions on Speech and Audio Processing, xx, 2000.

[11] Zhu, Q., Kneuer, J.G., Wei-Min Huang. (1995). A study of temporal behavior between the packet loss bursts in packet switched systems. Communications, 1995. ICC 95 Seattle, Gateway to Globalization, 1995 IEEE International Conference on , Volume: 3 , 18-22 June 1995. Page(s): 1930 -1936 vol.3

[12] Yoma, N.B., Busso, C., Soto, I. "Packet-loss modeling in IP networks with state duration constraint"s. Accepted for publication in IEE Proc. on Communications.